



## DATA MINING KLASIFIKASI KEPERIBADIAN SISWA SMP NEGERI 5 JEPARA MENGGUNAKAN METODE DECISION TREE ALGORITMA C4.5

Adjie Kukuh Wahyudi<sup>1</sup>, Noor Azizah<sup>2</sup>, Heru Saputro<sup>3</sup>

Universitas Nahdlatul Ulama Jepara

[kukuhwahyudi1234@gmail.com](mailto:kukuhwahyudi1234@gmail.com)<sup>1</sup>, [azizah@unisnu.ac.id](mailto:azizah@unisnu.ac.id)<sup>2</sup>, [herusaputro@unisnu.ac.id](mailto:herusaputro@unisnu.ac.id)<sup>3</sup>

### ABSTRACT

*In the world of education, there are many supporting factors to achieve success in the learning process, from teachers, materials, to students. The thing that still a problem for many educators is how to provide material or learning with various personality characteristics of students who arrive is still a difficulty for educators, this is what encourages the author to carry out this research, so that it can help educators to know the personality of each student. So research Data Mining Personality Classification of SMP Negeri 5 Jepara Students Using Decision Tree Algorithm C4.5 method. By using the decision method tree algorithm C4.5, this data mining can produce perfect output in classifying students' personalities, in this case study at SMP N 5 Jepara. These results are indicated by filling out questionnaires by researchers to 100 students and 1 teacher BP SMP Negeri 5 Jepara..*

**Keywords** : Data Mining, SMP N 5 Jepara, Algoritma C4.5.

### ABSTRAK

Didalam dunia pendidikan, banyak faktor penunjang untuk mencapai keberhasilan dalam proses pembelajaran, mulai dari guru, materi, sampai dengan siswa. Hal yang masih menjadi permasalahan banyak pendidik adalah bagaimana memberikan materi atau pembelajaran dengan berbagai karakter kepribadian siswa yang sampai saat ini masih menjadi kesulitan bagi para pendidik, hal inilah yang mendorong penulis untuk mengangkat penelitian ini, supaya dapat membantu para pendidik untuk mengetahui kepribadian masing-masing siswa. Maka dibuatlah penelitian Data Mining Klasifikasi Kepribadian Siswa Smp Negeri 5 Jepara Menggunakan Metode Decision Tree Algoritma C4.5. Dengan menggunakan metode decision tree algoritma C4.5, data mining ini dapat menghasilkan output yang sempurna dalam mengklasifikasi kepribadian siswa, dalam studi kasus ini di SMP N 5 Jepara. Hasil tersebut ditunjukkan dengan pengisian kuisioner oleh peneliti kepada 100 siswa dan 1 guru BP SMP Negeri 5 Jepara.

**Kata Kunci** : Data Mining, SMP N 5 Jepara, Algoritma C4.5.

### 1. PENDAHULUAN

Kepribadian yang dimiliki oleh siswa pada umumnya berbeda-beda, kepribadian yang dimiliki oleh setiap siswa akan mempengaruhi hasil belajar yang akan diperoleh oleh siswa pada saat proses pembelajaran berlangsung. Maka dari itu, dibutuhkan klasifikasi kepribadian masing-masing siswa. Proses klasifikasi dalam menentukan kepribadian siswa masih belum

optimal, serta dalam proses klasifikasi masih dilaksanakan secara manual belum adanya perhitungan akurat. Sehingga terdapat ketidaktepatan dalam klasifikasi kepribadian siswa. Berdasarkan permasalahan diatas penulis melakukan penelitian

mengenai klasifikasi kepribadian siswa Menggunakan data mining dengan metode decision tree. Data mining menunjukan sebagai

sistem yang bias memberikan data klasifikasi kepribadian siswa dengan lebih cepat dan akurat. Sistem ini menggunakan metode decision tree. Dengan menggunakan decision tree pada sistem dapat membantu memberikan solusi yang cepat dan tepat dalam menyajikan data klasifikasi kepribadian siswa.

Saat ini, klasifikasi kepribadian siswa di SMP N 5 Jepara masih dilakukan secara manual menggunakan kertas kuisioner, penilaian dilaksanakan secara manual satu persatu, yang kadang menghasilkan hasil yang kurang akurat, terlebih di masa pandemi seperti ini, tidak dapat melaksanakan klasifikasi, karena terkendala siswa tidak masuk sekolah, melaksanakan pembelajaran secara daring.

## 2. KERANGKA TEORI

### 2.1. Data Mining

Sebagai bidang ilmu yang relative baru, saat ini Data Mining menjadi salah satu pusat perhatian para akademis maupun praktisi. Menurut (Suntoro, 2019) data mining adalah proses untuk mendapatkan informasi yang berguna dari basis data yang besar dan perlu diekstraksi agar menjadi informasi baru dan dapat membantu dalam pengambilan keputusan. Data mining adalah proses menganalisa data dari yang berbeda dan menyimpulkannya menjadi informasi atau pengetahuan atau pola yang penting untuk meningkatkan keuntungan, memperkecil biaya pengeluaran, atau bahkan keduanya(Witten, 2016).

2.2. Karakteristik Utama Dan Tujuan Data Mining Berikut adalah karakteristik utama dan tujuan dari data mining:

1. Data seringkali terkubur dalam database yang sangat besar, yang terkadang berisi data selama bertahun-tahun. Dalam banyak kasus, data dibersihkan dan disatukan ke dalam data warehouse.
2. Environment data mining pada umumnya adalah arsitektur client-server atau arsitektur sistem informasi berbasis web.
3. Berbagai tool baru yang canggih, termasuk berbagai tool visualisasi yang canggih, membantu untuk mengangkat biji informasi yang terkubur dalam file file korporat atau record-record arsip. Untuk mendapatkannya akan melibatkan memoles dan mensinkronisasikan data untuk mendapatkan hasil-hasil yang tepat.

Data miners yang mutakhir juga memeriksa kemanfaatan data (misalnya, teks yang tak terstruktur yang disimpan dalam tempat-tempat seperti database Lotus Notes, file-file teks di internet, atau intranet korporat).

4. Si penambang seringkali adalah end-user, yang didukung dengan 'bor-bor data' dan berbagai tool query handal lainnya untuk menanyakan pertanyaan pertanyaan dengan tujuan tertentu dan mendapatkan jawaban-jawaban dengan cepat, dengan sedikit atau bahkan tanpa skill pemrograman sekalipun.

5. Dalam menemukan pola seringkali menemukan hasil yang tak diharapkan dan meminta end-user untuk berpikir secara kreatif dalam menjalankan proses, termasuk interpretasi terhadap temuan.

6. Banyak tool data mining siap dikombinasikan dengan berbagai spreadsheet dan tool development software lainnya. Jadi, data yang ditambang bisa dianalisa dan diterapkan dengan cepat dan mudah.

7. Karena jumlah data yang sangat besar dan usaha pencarian yang massif, kadang-kadang perlu menggunakan pemrosesan parallel untuk data mining

### 2.3. Pustaka Rujukan

Penelitian yang dilakukan penulis dengan beberapa tinjauan studi digunakan sebagai landasan teori dan pembanding dengan penelitian yang dilakukan. Sumber rujukan yang dijadikan acuan bagi penulis dalam melakukan penelitian ini diantaranya yaitu:

Pertama berjudul "Penerapan Data Mining Dalam Pengelompokkan Data Nilai Siswa Untuk Penentuan Jurusan Siswa Pada SMA Tamora Menggunakan Algoritma K-Means Clustering" diteliti oleh Yohanni Syahra, pada tahun 2018.

Data mining dengan algoritma K-Means dapat diterapkan pada SMA Tamora untuk menganalisis permasalahan yang ada yang berkenaan dengan pengelompokkan data nilai siswa untuk penentuan jurusan, aplikasi dirancang dengan mengadopsi algoritma K-Means yang dapat digunakan dalam pengelompokkan data nilai siswa untuk penentuan jurusan pada SMA Tamora, aplikasi yang dirancang dapat mengimplementasikan dan dapat dijadikan sebagai solusi pemecahan masalah dalam hal pengelompokkan data nilai

siswa untuk penentuan jurusan siswa pada SMA Tamora Penentuan Kriteria Meliputi 9 Kriteria yang masing-masing kriteria memiliki 3 subkriteria yaitu : Kepemilikan Telfon Seluler, Pekerjaan, Penghasilan,

Status Tempat Tinggal, Jenis Lantai, Kondisi Rumah, Fasilitas Jamban, Pendidikan, Wawancara Tetangga. Dari hasil penelitian ini diharapkan akan dapat membantu dan memudahkan pendamping PKH dalam proses seleksi penerima bantuan program keluarga harapan di Kabupaten Pesawaran Provinsi Lampung.

Kedua berjudul "Sistem Pakar Untuk Mengidentifikasi Kepribadian Siswa Menggunakan Metode Certainty Factor Dalam Mendukung Pendekatan Guru" diteliti oleh Nadya Andhika Putri, pada tahun 2018. Sistem Pakar untuk mengidentifikasi kepribadian siswa ini dimulai dengan pengumpulan data yang didapatkan dari seorang pakar, kemudian data tersebut akan dibuat menjadi basis pengetahuan untuk membangun Sistem Pakar mengidentifikasi Psikopat ini.

Ketiga berjudul "Klasifikasi Karakteristik Kepribadian Manusia Berdasarkan Tipologi Hippocrates-Galenus Menggunakan Metode Naïve Bayes" (Lolita Aprilla, 2017). Penelitian ini menerapkan data mining teknik klasifikasi dengan menggunakan metode Naïve Bayes untuk menentukan kelas klasifikasi karakteristik kepribadian siswa yaitu kelas sanguin, koleris, melankolis, atau plegmatis. Atribut yang digunakan pada penelitian ini terdiri dari 7 atribut, yaitu jenis kelamin, usia, sekolah, jawaban tes A, jawaban tes B, jawaban tes C, dan jawaban tes D.

Keempat berjudul "Identifikasi Kepribadian Siswa Melalui Penerapan Sistem Pakar Dengan Teknik Forward Chaining" (Tri Ginanjar Laksana, 2016). Aplikasi ini mampu mengidentifikasi kepribadian siswa serta dapat memberikan hasil kesimpulan berupa hasil diagnosa kepribadian berupa jenis kepribadian dan solusinya. Dengan adanya aplikasi ini dapat digunakan sebagai media pendekatan yang dilakukan siswa maupun guru bimbingan konseling.

Kelima berjudul "Sistem Pakar Penentuan Tipe Kepribadian Siswa Sekolah Dasar Menggunakan Metode Case Based Reasoning"

diteliti oleh Riko Hardianto, pada tahun 2018. Penentuan kepribadian siswa sekolah dasar dengan metode Case Based Reasoning adalah kegiatan untuk memetakan tingkah laku siswa di sekolah agar kenakalan-kenakalan siswa bisa diarahkan menjadi sesuatu yang bisa menunjang keberhasilan siswa dalam proses belajar dan mengajar. Dengan Case Based Reasoning adalah metode pencarian dalam sistem pakar yang biasa dan banyak digunakan perhitungan masalah psikologi. Pada metode ini terdapat 4 tahapan yakni Retrieve, Reuse, Revise, dan Retain. Karena sekolah dasar didaerah khusunya sangat jauh dari sekolah dasar di kota mengenai kepekaan kepekaan kepala sekolah mengevaluasi proses pembelajaran siswa dan pembelajaran yang diinginkan sesuai watak siswa. Berdasarkan beberapa penelitian tersebut yang membedakan dengan penelitian saat ini adalah metode penghitungan menggunakan metode decision tree algoritma C4.5.

#### 2.4. Decision Tree

Decision tree adalah salah satu metode klasifikasi yang paling populer, karena mudah untuk diinterpretasi oleh manusia. Decision tree adalah model prediksi menggunakan struktur pohon atau struktur berhirarki. Untuk memilih atribut akar, didasarkan pada nilai Gain tertinggi dari atribut-atribut yang ada.

Untuk mendapatkan nilai Gain, harus ditentukan terlebih dahulu nilai Entropy.

Rumus Entropy :

$$Entropy(s) = \sum_{i=1}^n -p_i * \log_2 p_i$$

$n$   
 $i=1$

S = Himpunan Kasus

n = Jumlah Partisi S

$p_i$  = Proporsi dari  $S_i$  dari terhadap S

#### 2.5. Lain-lain

Penggunaan Tanda numerik *Roman* untuk penomoran bab atau sub-bab adalah bersifat pilihan. Jika menggunakan tanda numerik *Roman*, maka bagian Pustaka Rujukan dan bagian ucapan terima kasih serta sub-judul atau sub-bab tidak ditandai dengan huruf. Gunakan dua spasi untuk memisahkan antar sub-bab. Gunakan tanda penghubung pada kata-kata yang dimodifikasi: "zero-field-cooled magnetitation", hindari susunan kalimat yang

tidak beraturan seperti, “Dengan menggunakan (1), beda potensial telah dihitung”, penulisan yang baik ialah “beda potensial dihitung dengan menggunakan persamaan (1),” atau “dengan persamaan (1), kita hitung beda potensial”.

Angka desimal gunakan angka nol sebelum titik untuk menuliskan angka desimal: “0,25”. Gunakan “cm<sup>3</sup>,” bukan “cc.” Tidak mencampurkan kata seutuhnya dengan singkatan dalam penulisan besaran fisis, misalnya: “weber/m<sup>2</sup>” melainkan “Wb/m<sup>2</sup>”. Gunakan kata seutuhnya ketika menulis unit satuan fisis dalam suatu kalimat: “beberapa henry..”.

### 3. Algoritma C4.5

Algoritma C4.5 merupakan algoritma yang digunakan untuk membentuk pohon keputusan Algoritma Decision. Secara umum algoritma C4.5 untuk membangun pohon keputusan adalah sebagai berikut:

1. Pilih atribut sebagai akar
2. Buat cabang untuk masing-masing nilai
3. Bagi kasus dalam cabang
4. Ulangi proses untuk masing-masing cabang sampai semua kasus pada cabang memiliki kelas yang sama
5. Untuk memilih atribut sebagai akar, didasarkan pada nilai gain tertinggi dari atribut-atribut yang ada.

Rumus Gain :

$$\text{Gain}(S, A) = \text{Entropy}(S) - \sum_{S_i} \frac{|S_i|}{|S|} \text{Entropy}(S_i)$$

S = Himpunan Kasus

A = Atribut

n = Jumlah Partisi Atribut

|S<sub>i</sub>| = Jumlah Kasus pada partisi ke-i

|S| = Jumlah Kasus dalam S

Contoh Kasus :

Tabel 2.1. Contoh Kasus

Nama Pertanyaan Jawaban

Adira Friza Agustin Saya orang yang menyukai pesta  
Kadang-kadang Saya orang yang mempunyai selera humor  
Selalu Saya suka menjadi pusat perhatian  
Tidak pernah

Penghitungan Total Entropy Node Jumlah Pertanyaan  
Tidak pernah (S1)

Kadang kadang (S2)

Selalu (S3)

Entropy Gain

1 Total 3 1 1 1 0,333 0,333

$\text{Entropy}(\text{Total}) = -1/3 * \log_2(1/3) + //$  node dibagi jumlah pertanyaan

$- 1/3 * \log_2(1/3) + //$  node dibagi jumlah pertanyaan

11

$- 1/3 * \log_2(1/3) //$  node dibagi jumlah pertanyaan

Penghitungan Gain

Gain Total = 0,333- (1/3\*0,333) = 0,222111 //

hasil penghitungan entropy

dikurangi node-entropy.

### 4. HASIL DAN PEMBAHASAN

Data mining kasifikasi kepribadian siswa SMP Negeri 5 Jepara telah melalui tahapan pengujian sistem dengan menggunakan metode blackbox testing. Hasil dari pengujian sistem tersebut sistem dapat berjalan dengan normal tanpa kendala dan bekerja sesuai dengan apa yang diharapkan ditandai pada saat sistem menerima inputan output yang dihasilkan oleh sistem telah sesuai dengan apa yang diharapkan. Semua menu bekerja dengan baik, dengan hasil tersebut maka dapat dikatakan bahwa tujuan pembuatan data mining kasifikasi kepribadian siswa SMP Negeri 5 Jepara ini sudah mencapai tujuan. Setelah sistem dibuat penulis juga melakukan uji coba sistem pada objek penelitian yaitu SMP Negeri 5 Jepara dengan cara menyebarkan kuesioner ke siswa SMP Negeri 5 Jepara. Dalam kegiatan ujicoba tersebut penulis memperoleh sekitar kurang lebih 100 responden dari siswa dan 1 responden dari admin, kemudian dilakukan perhitungan analisa hasil kuesioner. Rata rata yang didapat dari perhitungan tersebut adalah 75,88% dari hasil pengujian sistem sehingga menunjukkan hasil yang baik. Diharapkan dengan adanya sistem tersebut dapat diterapkan pada sekolah yang lain pada bagian bimbingan konseling masih serba manual dalam klasifikasi kepribadian siswa sehingga dapat membantu admin dan guru BP dalam klasifikasi kepribadian siswa.

### 5. SIMPULAN

Berdasarkan penelitian yang dilakukan oleh penulis tentang data mining kasifikasi kepribadian siswa SMP Negeri 5 Jepara maka, dapat disimpulkan bahwa :

:

1. Implementasi Data Mining dengan metode decision tree dan algoritma C4.5 Pada data mining klasifikasi kepribadian siswa dapat memudahkan guru BP dalam pengklasifikasian kepribadian siswa.

2. Data mining klasifikasi kepribadian siswa dapat menghasilkan output yang akurat.
3. Hasil kuisioner pengujian Data Mining dengan metode decision tree dengan algoritma C4.5 dari 100 responden siswa dan 1 guru BP SMP Negeri 5 Jepara, dapat diambil hasil besar bahwa sistem ini dapat berjalan dan berfungsi dengan baik sebagaimana mestinya.

#### UCAPAN TERIMAKASIH

Terimakasih kepada seluruh keluarga Universitas Islam Nahdlatul Ulama Jepara, SMP Negeri 5 Jepara.

#### DAFTAR PUSTAKA

- Yohanni Syahra. (2018). Penerapan Data Mining Dalam Pengelompokan Data Nilai Siswa Untuk Penentuan Jurusan Siswa Pada SMA Tamora Menggunakan Algoritma K-Means Clustering, E-ISSN : 2615-3475.
- Nadya Andhika Putri. (2018). Sistem Pakar Untuk Mengidentifikasi Kepribadian Siswa Menggunakan Metode Certainty Factor Dalam Mendukung Pendekatan Guru. *Journal of Information Technology and Computer Science* 1(1): 78-90.
- Wildan Hakim, Turmudi, Wahyu H. Irawan. (2015). Sistem Pendukung Keputusan Metode Sugeno dalam Menentukan Tingkat Kepribadian Siswa Berdasarkan Pendidikan (Studi Kasus di MI Miftahul Ulum Gondanglegi Malang).
- Didik Setiyadi, Ali Nurdin. (2012). Data Mining Potensi Akademik Siswa Berbasis Online. Didik Setiyadi, Ali Nurdin.
- Imam Sutoyo. 2018. Implementasi Algoritma Decision Tree Untuk Klasifikasi Data Peserta Didik. *Jurnal PILAR Nusa Mandiri* Vol. 14
- Kusrini dan Emha Taufiq Luthfi. 2009. Algoritma Data Mining. Penerbit, Yogyakarta.
- Dewi, M.R., Mudakir, I. and Murdiyah, S. Pengaruh Model Pembelajaran Kolaboratif berbasis Lesson Study terhadap Kemampuan Berpikir Kritis Siswa. *Jurnal Edukasi*, 3 (2), pp.29-33, 2016.
- B. Liu. Sentiment Analysis and Opinion Mining. USA: Morgan &
- Mello, R.F. and Gašević, D., 2019, October. What is the effect of a dominant code in an epistemic network analysis?. In *International Conference on Quantitative Ethnography* (pp. 66-76). Springer, Cham.
- Shaffer, D.W., Collier, W. and Ruis, A.R. A tutorial on epistemic network analysis: Analyzing the structure of connections in cognitive, social, and interaction data. *Journal of Learning Analytics*, 3 (3), pp.9-45, 2016.
- Breiman, L.. Random forests. *Machine learning*, 45 (1), pp.5-32, 2001.
- Arifin, O. and Sasongko, T.B. Analisa Perbandingan Tingkat Performansi Metode Support Vector Machine dan Naïve Bayes Classifier Untuk Klasifikasi Jalur Minat SMA. *SEMNAS TEKNO MEDIA ONLINE*, 6 (1), pp.1-2, 2018.
- Sari, A.P., Saptono, R. and Suryani, E. The Implementation of Jaro Winkler Distance and Naive Bayes Classifier for Identification System of Pests and Diseases on Paddy. *ITSMART: Jurnal Teknologi dan Informasi*, 7 (1), pp.1-7, 2018.
- Lestari, A.R.T., Perdana, R.S. and Fauzi, M.A. Analisis Sentimen Tentang Opini Pilkada DKI 2017 pada Dokumen Twitter Berbahasa Indonesia Menggunakan Naive Bayes dan Pembobotan Emoji. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer* e-ISSN, 2548, p. 964X. 2017.
- Ling, J., Kencana, I.P.E.N. and Oka, T.B. Analisis Sentimen Menggunakan Metode Naïve Bayes Classifier dengan Seleksi Fitur Chi Square. *E Jurnal Matematika*, 3(3), pp.92-99, 2014.
- Sari, R. and Hayuningtyas, R.Y. Penerapan Algoritma Naive Bayes Untuk Analisis Sentimen Pada Wisata TMII Berbasis Website. *Indonesian Journal on Software Engineering (IJSE)*, 5(2), pp.51-60, 2019.
- Marquart, C.L. (2019). rENA: Epistemic Network Analysis. [ONLINE] Available at: <https://cran.rproject.org/web/packages/rENA/index.html>.
- Leydesdorff, L. and Welbers, K. The semantic mapping of words and co-words in contexts. *Journal of Informetrics*, 5(3), pp.469-475, 2011.
- Swiecki, Z. and Shaffer, D.W. iSENS: an integrated approach to combining epistemic and social network analyses. In *Proceedings of the Tenth International Conference on*

Learning Analytics & Knowledge (pp. 305-313), 2020.